REMARKS

This application has been reviewed in light of the Office Action dated November 22, 2005. In view of the foregoing amendments and the following remarks, favorable reconsideration and withdrawal of the rejections set forth in the Office Action are respectfully requested.

Claims 1, 6, 7, 9-13, 18, 19 and 21-33 are pending. Claims 1, 9, 13, 21, 25 and 28 have been amended. Claims 32 and 33 have been added. Support for the new claims and claim changes can be found in the original disclosure, and therefore no new matter has been added. Claims 1, 13, 25, 32 and 33 are in independent form.

Claims 1, 12, 13, 24, 25 and 31 were rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,101,470 (*Eide et al.*).

Claims 6, 7, 18, 19, 26 and 27 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Eide et al.* in view of U.S. Patent No. 5,913,193 (*Huang et al.*).

Claims 9, 21 and 28 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Eide et al.* in view of U.S. Patent No. 6,161,091 (*Akamine et al.*).

Claims 10, 11, 22, 23, 29 and 30 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Eide et al.* in view of U.S. Patent No. 6,665,641 (*Coorman et al.*)

Without conceding the propriety of the rejections, independent Claims 1, 13 and 25 have been amended herein. Applicants submit that independent Claims 1, 13 and 25, as amended, and newly presented independent Claims 32 and 33 are patentable over the applied art for at least the following reasons.

Independent Claim 1 is directed to a speech signal processing apparatus comprising obtaining means for obtaining a plurality of synthesis units based on an input text, modifying means for modifying each of the synthesis units according to prosody information obtained based on the input text, distortion obtaining means for obtaining a respective modification distortion for each of the plurality of synthesis units, based on the respective synthesis unit before modification and that synthesis unit after modification, selection means for selecting synthesis units based on the modification distortions obtained by said distortion obtaining means, and speech synthesis means for performing speech synthesis based on the synthesis units selected by said selection means. Each of independent Claims 13, 25, 32 and 33 recites identical or similar features.

Applicants submit that nothing in the applied art would teach or suggest at least the distortion obtaining means for obtaining a respective modification distortion for each of the plurality of synthesis units, based on the respective synthesis unit before modification and that synthesis unit after modification, or the selection means for selecting synthesis units based on the modification distortions obtained by said distortion obtaining means, as recited in the claimed combination of independent Claim 1.

For a description of an example of the recited modification distortion, the Examiner is respectfully directed to Applicants' specification at paragraphs [0047] and [0048], with reference to Fig. 7.

Eide et al. relates to methods for generating pitch and duration contours in a text to speech system. According to *Eide et al.*'s invention, input text is converted to an output acoustic signal simulating natural speech, using a method that models certain prosodic features to

improve the naturalness of the synthetic speech produced. A text is input, and its lexical stress levels are calculated. Then, the calculated stress levels are compared to stored training data to find the training data that most closely matches the input text. When the most closely matching training data has been found, the pitch levels associated with the stress levels thereof are copied to generate a pitch contour to be assigned to the input text.

More specifically, *Eide et al.* operates as follows. A database or corpus of training sentences is obtained from one or more speakers. For each training sentence, each word is broken down into its constituent phonemes according to a phonetic dictionary, and the stress levels of the words forming the sentence are concatenated to obtain a stress contour for the sentence. (Each entry in the phonetic dictionary consists of a sequence of phonemes forming a word, and a sequence of stress levels corresponding to the vowels in the word.) Each stress level is associated with a pitch level, in a stress and pitch level pair. A pitch contour is obtained for each training sentence by concatenating the pitch levels associated with the stress levels. For each training sentence, a sequence of its stress and pitch level pairs is stored. The collection of all such sequences (corresponding to all the training sentences, respectively) is referred to as a pitch model.

In speech synthesis, the stress contour of the input sentence is calculated, in the same manner as was done for the training sentences. Then, the end of the stress contour of the input sentence is aligned with the ends of the stress contours of the training sentences. Then, each of the aligned stress contours is segmented, from its end to its beginning, into a plurality of (stress contour) blocks. Thus, each block of the input sentence is aligned with a number of training blocks equal to the number of training sentences used to create the pitch model.

Then, starting with the last input block (i.e., the block at the end of the stress contour of the input sentence), its stress contour (stress levels) is compared with that of each training block that is aligned with that input block (i.e., compared with the last block of each training sentence). The training block whose stress contour most closely matches that of the input block is selected. This process is repeated for the penultimate input block: its stress contour is compared with that of each training block aligned therewith (i.e., each penultimate training block), and the training block with the most closely matching stress contour is selected. This process is repeated for each of the remaining input blocks, until the beginning of the input sentence is reached.

As a result of performing this comparison and selection process for each of the input blocks of the input sentence, there is obtained a sequence consisting of the training blocks that were selected as having the most closely matching stress contours. The pitch values of that sequence (i.e., the pitch values associated with the stress levels of the sequence) are concatenated to form a pitch contour that is (subject to further modification) to be applied to synthesis units to create synthesized speech of the input sentence.

It is noted that the training blocks selected as most closely matching may not be exact matches but may include discrepancies with the corresponding input blocks (i.e., mismatched stress levels) at particular positions. Accordingly, in the pitch contour to be used to create synthesized speech of the input sentence, the pitch values at such discrepant positions may be adjusted to reduce the magnitude of the discrepancies.

In addition, the durations of the phonemes of the waveform segments to be output as synthesized speech are adjusted in accordance with their stress levels.

In sum, *Eide et al.*'s speech synthesis operates as follows. For each input text sentence to be synthesized, a pitch contour is created by concatenating units (blocks) of training data whose stress contours most closely match those of corresponding blocks of the input text sentence. The pitch contour is also adjusted to compensate for residual discrepancies between the most closely matching training data and the input text. The durations of the phonemes of the input text sentence are also adjusted according to their stress levels. Speech synthesis is performed using the pitches and durations thus generated.

The Office Action (page 3) cites col. 3, lines 35-45, col. 4, lines 32-40, col. 5, lines 9-26 and col. 8, lines 42-53 of *Eide et al.* as teaching "[d]istortion obtaining means for obtaining a modification distortion between synthesis units before and after modification responsive to prosody of a text." However, even if the cited portions of *Eide et al.* be deemed to teach that synthesis units of an input text are modified with respect to their prosody, Applicants submit that they do not teach or suggest that a modification distortion, based on a synthesis unit before modification and that unit after modification, is obtained, as recited in independent Claim 1.

Col. 3, lines 35-45 of *Eide et al.* teaches that input text is segmented into a sequence of phonemes, and mapped to a sequence of stress levels. Then, a waveform segment is selected for each phoneme. Pitch and duration contours are selected for the sequence of phonemes. The selected waveform segments are combined and their pitch and durations are adjusted to generate the output acoustic signal simulating natural speech. Even if, for the sake of argument, this portion of *Eide et al.* be deemed to teach that speech synthesis units (waveform segments) are selected and that the synthesis units are modified (pitch and durations are

adjusted), Applicants submit that it does not suggest obtaining a modification distortion between a synthesis unit before and after modification (or selecting synthesis units based on such modification distortions).

Col. 4, lines 32-40 of *Eide et al.* teaches that after training of the system, the input text to be synthesized is obtained, and each input sentence thereof is matched to an utterance type (e.g., declaration, question, exclamation). Then, the stress contour of each input sentence is calculated, which involves expanding each word of the sentence into its constituent phonemes according to a phonetic dictionary and concatenating the stress levels of the words in the dictionary forming the sentence. Thus, this portion of *Eide et al.* teaches that the stress contour of each input sentence is calculated. (As was explained above, this is a step toward finding a stress contour in the training data that most closely matches that of the input sentence, after which the pitch contour associated with the most closely matching stress contour will be used in creating synthetic speech of the input sentence.) Even if, for the sake of argument, this portion of *Eide et al.* be deemed to teach a step in a process of obtaining and modifying speech synthesis units, Applicants submit that this portion of *Eide et al.* does not suggest obtaining a modification distortion between a synthesis unit before and after modification (or selecting synthesis units based on such modification distortions).

Col. 5, lines 9-26 of *Eide et al.* teaches that stress levels of input and training sentences are to be compared. In order to do this, the stress contours of the input and training sentences are segmented, which involves aligning the ends of the stress contours of the input and training sentences, and respectively segmenting the stress contours from the ends towards the beginnings of the sentences. The result of the segmentation is a plurality of input stress contour

blocks respectively aligned with a plurality of training stress contour blocks. For every input block, there will be a corresponding number of aligned training blocks. Thus, this portion of *Eide et al.* teaches segmenting and aligning the stress contour of an input sentence with those of the training sentences. (As was explained above, this too is a step toward finding the stress contour in the training data that most closely matches that of the input sentence, after which the pitch contour associated with the most closely matching stress contour will be used in creating synthetic speech of the input sentence.) Even if, for the sake of argument, this portion of *Eide et al.* be deemed to teach a step in a process of obtaining and modifying speech synthesis units, Applicants submit that this portion of *Eide et al.* does not suggest obtaining a modification distortion between a synthesis unit before and after modification (or selecting synthesis units based on such modification distortions).

Col. 8, lines 42-53 of *Eide et al.* teaches that the best pitch contour to be used for synthesizing a given input sentence is obtained by comparing, in blocks, the stress contours of the input sentence to the stress contours of the training sentences, in order to find the training stress contour blocks that most closely match the input stress contour blocks. The closest match is found by computing, for each input block, the distance from the input block to each training block aligned therewith. Here, the Euclidean distance is measured, but it is noted that the selection of a distance measure is arbitrary. Thus, this portion of *Eide et al.* teaches comparing the stress contours of the input and training sentences (blocks) to find the training stress contour blocks most closely matching a given input sentence, by measuring the distance between the input and training blocks. This is done in order to obtain the best pitch contour to be used in synthesizing speech of the input sentence. (As was explained above, this portion is teaching

finding the stress contour in the training data that most closely matches that of the input sentence (by computing the distance between the input and training blocks), for the purpose of obtaining the pitch contour associated with the most closely matching stress contour, which will be used in creating synthetic speech of the input sentence.) Even if, for the sake of argument, this portion of *Eide et al.* be deemed to teach a step in a process of obtaining and modifying speech synthesis units, Applicants submit that this portion of *Eide et al.* does not suggest obtaining a modification distortion between a synthesis unit before and after modification (or selecting synthesis units based on such modification distortions).

The Office Action (page 3) cites col. 5, lines 27-48 and col. 8, lines 42-53 of *Eide et al.* as teaching "[s]election means for selecting synthesis units based on the modification obtained by said distortion obtaining means." However, even if the cited portions of *Eide et al.* be deemed to teach that synthesis units of an input text are modified with respect to their prosody, Applicants submit that they do not teach or suggest selecting synthesis units based on modification distortions obtained by a distortion obtaining means, as recited in independent Claim 1.

Col. 5, lines 27-48 of *Eide et al.* teaches that, for each input sentence, the stress levels of each input block are compared to the stress levels of the corresponding aligned training blocks in order to obtain a sequence of training blocks having the stress levels closest to those of the input blocks. It is noted that each stress level in a training block is associated with a pitch level in a stress and pitch level pair. When a closest sequence of training blocks is obtained for an input sentence, the pitch levels associated with the stress levels of the obtained sequence are concatenated to form a pitch contour for that input sentence. In addition, the durations of the

phonemes of the words of the input sentences are adjusted based on the stress levels. In addition, pitch levels of the pitch contours may be adjusted if their associated stress levels do not match the corresponding stress levels of the corresponding input blocks. Thus, this portion of *Eide et al.* again teaches comparing input and training stress levels to find the training stress contour most closely matching the input stress contour, and then forming a pitch contour from the pitch levels associated with the stress levels of the most closely matching training stress contour. This portion of *Eide et al.* also teaches adjusting the durations of phonemes based on stress levels and adjusting the pitch levels of the pitch contours if the training stress level does not match the corresponding input stress level. *Eide et al.* is not understood to suggest that selection of synthesis units occurs after the assigning of the pitch levels to synthesis units, after the adjustment of the durations, or after the adjustment of the pitch levels, based on those respective assignations/adjustments. Even if this portion of *Eide et al.* be deemed to teach or suggest modifying speech synthesis units (e.g., assigning pitch levels to synthesis units, adjusting durations of synthesis units, or adjusting pitch levels of synthesis units), Applicants submit that this portion of *Eide et al.* does not suggest (obtaining a modification distortion between a synthesis unit before and after modification, or) selecting synthesis units based on such modification distortions.

Col. 8, lines 42-53 of *Eide et al.* has been discussed above. As discussed, this portion of *Eide et al.* is submitted not to teach or suggest (obtaining a modification distortion between a synthesis unit before and after modification, or) selecting synthesis units based on modification distortions.

In view of the explanation of *Eide et al.* given above, Applicants submit that no portion of *Eide et al.* other than the specific portions cited by the Office Action discussed above would teach or suggest obtaining a modification distortion between a synthesis unit before and after modification, or selecting synthesis units based on such modification distortions.

Since *Eide et al.* is not understood to teach all of the elements of independent Claim 1, that claim is believed allowable over that document. Since each of independent Claims 13, 25, 32 and 33 recite features identical or similar to those of Claim 1, those claims are believed allowable over *Eide et al.* for at least the same reasons.

(Regarding independent Claims 32 and 33, it is noted that they do not recite obtaining a "modification distortion . . ., based on the respective synthesis unit before modification and that synthesis unit after modification," but that they recite obtaining a "modification distortion . . . from a modification distortions table according to prosody information obtained based on the input text, the modification distortions corresponding to the synthesis units in the database, respectively, and being obtained by modifying the synthesis units based on prosody information set in the modification distortions table." Nonetheless, since it has been argued above that *Eide et al.* does not teach or suggest obtaining a modification distortion between a synthesis unit before and after modification, or selecting synthesis units based on such modification distortions, the arguments set forth above are understood to apply to independent Claims 32 and 33.)

(If, contrary to Applicants' arguments set forth above, the Examiner still believes that *Eide et al.* teaches obtaining a modification distortion between a synthesis unit before and after modification, or selecting synthesis units based on such modification distortions,

the Examiner is respectfully requested to identify specifically which portions of *Eide et al.* are deemed to teach these features and the manner in which they are deemed to teach them.)

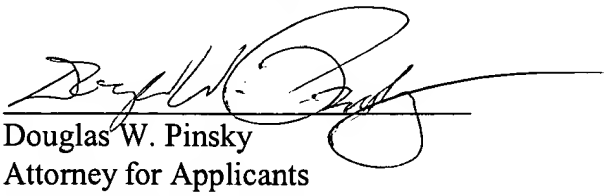
A review of the other art of record, including *Huang et al.*, *Akamine et al.* and *Coorman et al.*, has failed to reveal anything which, in Applicants' opinion, would remedy the deficiencies of *Eide et al.*, as a reference against the independent claims herein. These claims are therefore believed patentable over the art of record.

The other claims in this application are each dependent from one or another of the independent claims discussed above and are therefore believed patentable for the same reasons. Since each dependent claim is also deemed to define an additional aspect of the invention, however, the individual reconsideration of the patentability of each on its own merits is respectfully requested.

In view of the foregoing amendments and remarks, Applicants respectfully request favorable reconsideration and early passage to issue of the present application.

Applicants' undersigned attorney may be reached in our Washington office by telephone at (202) 530-1010. All correspondence should continue to be directed to our below listed address.

Respectfully submitted,



Douglas W. Pinsky
Attorney for Applicants
Registration No. 46,994

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200
DWP/klm

DC_MAIN 231874v1